Introduction

The 1996 Health Care Survey of DoD Beneficiaries (HCSDB) is the second annual survey conducted by the Department of Defense (DoD). DoD is required by law to conduct this annual evaluation of the health care services provided by the Military Health Care System. The 1996 HCSDB investigated beneficiaries' opinions concerning their access to, familiarity and satisfaction with the Military Health Care System. Questions on the health status of beneficiaries and attitudes toward Tricare were also included in the survey.

The purpose of this Codebook is to document the data collection activities and survey database elements for the 1996 HCSDB Form A (adult beneficiaries). Its primary goal is to document data collection activities during the field period, and to present information on each variable in the survey data file. A detailed discussion of how variables were coded is presented in various sections of the codebook.

This codebook contains the following information:

- Overview of how beneficiaries were selected for the survey.

- Information on the survey administration cycle, with specific details on the survey mailing cycle.

- Information about the number of surveys received, and the decision rules used to remove records from the survey file.

- An annotated questionnaire that provides the names of survey variables, and their respective values.

- Information on how survey variables were coded and recoded based on skip patterns within the questionnaire.

- Details on each variable in the data base, including frequencies of coded values, missing values, and not applicable values.

- A list of variables in the final data file generated from a SAS data set.

- A record layout for use in reading the data file in flat file format.

- A crosswalk of questions from the 1996 and the 1994-95 survey.

Overview of Sampling Design

The sample for the 1996 HCSDB (Form A) was supplied by DMDC from the Defense Enrollment Eligibility Reporting System (DEERS) files. The sample was stratified by 149 catchment areas (CACSMPL) and six beneficiary groups (BGCSMPL). The catchment area strata included catchment areas around military hospitals, some catchment areas around clinics, and areas defined as "non-catchment" areas outside hospital catchment areas. The beneficiary groups eligible for participation in the survey included: active duty personnel, family members of active duty personnel,

retirees under age 65; retirees age 65 or over; family members age 18-64 of retirees and survivors age 18-64; and family members age 65 or over of retirees and survivors age 65 or over.

The sample of beneficiaries selected for the survey was based on current eligibility information as of October 28, 1995. The 1996 HCSDB sample of 156,838 beneficiaries was selected from a sampling frame of 6,455,915 eligible beneficiaries. A detailed discussion of the sampling design and the sampling approach can be found in Chu, Flores-Cervantes, and Latta (1996). Table 1 provides a frequency distribution of sample members selected for the survey by beneficiary group.

Table 1.
Frequency distribution of 1996 HCSDB sample members by beneficiary groups

Beneficiary Group	Frequency (n)	Percent of Sample
Active Duty Personnel	38,214	24.4
Family Members of Active Duty Personnel	30,725	19.6
Retirees Under Age 65	22,205	14.2
Retirees Age 65 or Over	17,145	10.9
Family Members Age 18-64 of Retirees and Survivors Age 18-64	29,213	18.6
Family Members Age 65 or Over of Retirees and Survivors Age 65 or Over	19,336	12.3
Total	156,838	100.0

Survey Operations Activities

The operational support for mailing the survey involved four mailings to beneficiaries between April 15, 1996 and August 8, 1996. The mailings were scheduled to maximize response rates for the data collection effort. Targeted mailings and remailings have been demonstrated to increase response rates. First, an initial notification letter was mailed to the entire sample. Four weeks after the start date of the initial notification mailing, a first-wave of surveys was mailed to the entire sample. Approximately two weeks after the first wave of survey mailings, a thank-you/reminder letter was mailed. Finally, four weeks after the mailing of the follow-up letters, a second wave mailing of surveys was directed to those who had failed to respond to previous mailings. The field was closed for targeted mailings on September 10, 1996.

Address Update Activities Prior to and During Survey Administration

Upon receipt of the sample file from DMDC, the addresses were examined to determine whether an address was suitable for mailing. Within each record, a priority was assigned to each address based on its source and type, e.g., DEERS residence address, DEERS unit address. All sample records (excluding foreign countries) with sufficient address information were sent to an outside vendor where they were interfaced with the National Change of Address (NCOA) database to obtain updated address information. Addresses outside the U. S. were not submitted, as they are not included in the NCOA database. The NCOA update was done twice; 146,587 records were sent prior to the notification letter mailing and 149,095 records were sent prior to the Wave 1 mailing. The second submission contained a higher number of records because additional addresses were obtained from commercial credit bureau sources (Equifax, TransUnion, TRW). Each time, NCOA returned the updated address file to the operations contractor and the updated NCOA address file was integrated with the DMDC-provided data in the system used for mailing. In the notification letter mailing, the NCOA-provided address was labeled as the highest priority address in the system file and was the first address attempted. The highest priority address for each record was selected. Next, records selected for a major mailing (excluding remails) were sorted according to first class presort postal regulations using Group 1 software[1]. A print file[2] was then created, to be used in producing the personalized cover letters.

The updating of addresses was a continuous process throughout the survey administration cycle. During survey administration, address updates were obtained in multiple ways: when sample members self-reported information via telephone (using the 800-number system designated for calls regarding this survey), fax, or letter; when the postal service forwarded address correction information (ACRs); when the postal service returned letters or packets with out-of-date forwarding (ODFs) but with new address information affixed to the envelopes; or when letters or packets were returned as postal non-deliverables (PNDs). To obtain new address information for PNDs, these records were submitted to three commercial credit bureaus simultaneously.

Address information received directly from a beneficiary was considered the most accurate and received the highest address priority. The notification and reminder letters included a toll-free telephone number and numbers for faxes and collect calls (for non-U.S. sample members), so that beneficiaries would be aware of an easy and free method of updating their own addresses as necessary. Next highest priority was address information received from the post office in response to the "Address Correction Requested" legend printed on the carrier envelopes. This consisted of a photocopy of the forwarded envelope with the change of address information noted. This information was from the post office's database of address correction cards filed by people who have moved.

When a letter or survey was returned PND, the associated record was labeled to reflect that the survey packet was returned PND and that the address was invalid and therefore unusable. The record is then flagged for inclusion in the next remail. The next-in-line address was identified

---

[1] Group 1 software standardizes addresses according to postal requirements.
[2] The print file is the file of names and addresses to be printed on the cover letters.

for use in the next mailing. Each address within a given record was used based on its assigned priority. Once all addresses had been used, the record was flagged for inclusion in the next submission to the credit bureaus, prior to the next remail. Submission to the credit bureau was a final attempt to obtain updated address information.

Out of the total sample, 10,251 sample members had insufficient address information in the address fields (for all available addresses). Any record without a usable address was sent to the three commercial credit bureaus for an address search. The credit bureaus returned all records to the operations contractor with updated address information, if available. (Credit bureau updates included the receipt date of new address information as part of the record returned to the operations contractor, which allowed them to be prioritized by the most recent date received.) The updates were added to the mailing file and were labeled as the highest priority addresses. Where multiple addresses were received, the most recent address was chosen as highest priority, with others as the next-in-line addresses. The mailing of letters or surveys to these sample members was then conducted, following the same steps as the original mailing.

Credit bureau submissions occurred subsequent to receipt of the NCOA updates, as records that NCOA could not update were sent to the credit bureaus for address searches. Also, in accordance with the contract requirements, records for which the address was identified as a PND or ODF were submitted to the credit bureaus approximately every three days on an ongoing basis throughout the survey administration cycle. Any PNDs received after the cut-off date were processed in the next batch sent to the credit bureau.

The following is a summary of the order in which the addresses were prioritized in the mailing system, from highest to lowest priority:

1. Contact from sample member (phone call, voice mail, fax, letter, returned survey)
2. Update from post office (ACRs, ODFs)
3. Update from NCOA
4. Update from commercial credit bureaus
5. DEERS residence address
6. DEERS unit address

Table 2 summarizes the final address sources for surveys included in the 1996 HCSDB data file. The table shows that only about one-fourth of the final data set consisted of surveys from updated sources such as the 800-number system, NCOA, and commercial credit bureaus.

Table 2.
Frequency of Address Sources for 1996 HCSDB Data File

Address Type	Frequency (n)	Percent of Returns
DEERS residence	68,459	76.3
DEERS unit address	3,078	3.4
800-number information	1,563	1.7
Fax or mail	404	.5
NCOA data base	6,184	6.9
Commercial Credit Bureau (Equifax)	993	1.1
Commercial Credit Bureau (TransUnion)	1,962	2.2
Commercial Credit Bureau (TRW)	4,551	5.1
Address Correction Requests from Postal Service	2,507	2.8

Letter Processing Procedures

Mailings which did not include a survey were generated and printed with the "best available" address from the system used for mailing. This address may have been the address generated from the DEERS file, NCOA, or commercial sources (Equifax, TransUnion, or TRW) or obtained through contact with the sample member (telephone, letter or fax) or from the postal service (address corrections). Each letter was printed with a unique identifier in the address block and the lower right corner, so that the beneficiary could refer to the number if address corrections were requested by fax or phone. Letters and packets with surveys were sent by first class mail.

The procedure for mailing surveys was more complex. Prior to the production of letters, each record in the mailing was matched with an available survey identification number (survey ID). As each survey ID was assigned, it was also recorded in the system used for mailing. Cover letters printed with each sample member's assigned survey ID were generated and printed in survey ID order. The letters were paired with the matching survey lithocode[3], inserted into envelopes with postage-paid return envelopes enclosed and sent by first class mail. A 10-percent quality control check was implemented to ensure that the surveys and letters contained the same survey ID. If an error was found, the packets were opened, examined, and the correct survey ID lithocode combination was made.

---

[3] Lithocodes are the survey identification numbers printed on the survey questionnaires in a binary format, so that they can be read by the OMR scanner and converted into Arabic numbers for the data file.

Survey Administration Time Line

The HCSDB mailing process was designed so that each sample member with a usable address was sent up to four documents: a notification letter, a first wave survey, a reminder/thank-you letter, and a second wave survey. If a sample member returned a survey during the first wave mailing, then a second wave survey was not sent. If a sample member was identified as deceased, that record was updated as such and no longer included in the mailing process. In the mailing process described below and in Table 3, the dates cited are the dates on which records for the mailing were selected and a print file created. The packets were usually mailed from one to five days after the print file was created.

The notification letter mailing, for which the print file was created on April 15, 1996, consisted of 146,502 letters sent to all sample members except those who had no known address. Those records were subsequently sent to credit bureaus (10,251). A small number of special cases, which consisted of records with addresses that had been rejected by the Group 1 software and records for people who were transitioning out of the military, were included in later mailings. The notification letters were sent to notify the beneficiaries that they were selected for the survey and to provide information to the beneficiaries regarding address updating procedures if the letters had been forwarded or had incorrect addresses. There were five remailings of the notification letter to sample members where the notification letters were returned as postal non-deliverables or for sample members previously without a known address where an address resulted from the credit bureau search. These remailings took place between April 20 and June 17, 1996 and totaled 24,810 pieces.

The first wave survey mailing, for which the print file was created on May 15, 1996, consisted of 144,213 surveys sent to all sample members except those who had not yet been mailed the notification letter. A print file for an additional Wave 1 mailing to 11,650 sample members was created on May 30, 1996. This group was delayed from the initial Wave 1 mailing to allow for a mailing of the notification letter. For Wave 1, each sample member received a survey, a cover letter requesting that the sample member complete and return the survey, and a return envelope. Three remailings totaling 4,364 packets were targeted to sample members whose surveys were returned as postal non-deliverable. These remailings took place between June 7 and June 17, 1996.

The reminder/thank-you letter mailing (for which the print file was created on May 29, 30 and 31, 1996) consisted of 143,275 letters sent to all sample members, with the exception of those who had not been mailed the Wave 1 packets. A print file for an additional reminder letter mailing, consisting of 13,042 packets, was created on June 7, 1996. The reminder/thank-you letter was sent to thank the sample member for completing the survey and to encourage the sample member to return the survey if one had not been completed. The reminder/thank-you letter also contained address updating procedures if the letter had been forwarded or had an incorrect address. Seven remailings were targeted to sample members whose letters were returned as postal non-deliverables. These remailings took place between June 13 and July 31, 1996 and totaled 8,162 pieces.

The second wave mailing (for which the print file was created on June 27, 1996) consisted of 86,627 surveys sent to sample members who had not completed and returned a survey. Excluded from this group were active refusers (those who made a verbal or written request to not participate), deceased beneficiaries, and beneficiaries who were ill, hospitalized, or incarcerated[4]. One remailing of 4,173 pieces was created on August 8, 1996.

Table 3 summarizes the various HCSDB mailings as recorded in the system used for mailing. The data includes the type of mailing; the date(s) the records were selected for inclusion in the mailing; and the number of letters sent.

Table 3.
Mailing time line

Mailing Type	Date of Selection	Number Mailed
Notification Letter	4/15/96	146,502
Notification Letter Remail	4/20/96	9,673
Wave 1 Survey	5/15/96-5/16/96	144,213
Notification Letter Remail	5/29/96	11,647
Wave 1 (additional mailing)*	5/30/96	11,650
Reminder Letter	5/29/96-5/31/96	143,275
Notification Letter Remail	6/7/96	1,321
Wave 1 Remail	6/7/96	1,549
Reminder Letter (additional mailing)*	6/7/96	13,042
Notification Letter Remail	6/13/96	887
Wave 1 Remail	6/13/96	1,378
Reminder Letter Remail	6/13/96	962
Notification Letter Remail	6/17/96	1,282
Wave 1 Remail	6/17/96	1,437
Reminder Letter Remail	6/18/96	1,300
Reminder Letter Remail	6/20/96	1,142

---

[4] Active refusers and sample members who were found to be ill, hospitalized, or incarcerated were sent a Wave 1 Survey; however, they did not receive the Wave 2 Survey.

Table 3. (continued)
Mailing time line

Mailing Type	Date of Selection	Number Mailed
Wave 2 Survey	6/27/96	86,627
Reminder Letter Remail	7/2/96	2,440
Reminder Letter Remail	7/10/96	464
Reminder Letter Remail	7/25/96	1,370
Reminder Letter Remail	7/31/96	484
Wave 2 Remail	8/8/96	4,173

* These records were not received from the credit bureaus in time to be included in the Wave 1 or Reminder Letter mailing. They were inserted and mailed after receipt of updated address information.

Processing of Incoming Surveys

Incoming survey forms were visually checked prior to scanning. Blank forms or partially completed forms with notations regarding the reason for returning the survey were divided into batches according to the reason the sample member wrote on the returned form. A respondent's reason for returning a blank or partially completed form was recorded in the mailing system. Surveys were then optically scanned so that lithocodes could be captured and tracked. This tracking of survey IDs was used to identify whether a sample member had returned a survey or not and to record the reason given for a blank return.

Blank forms without an explanation for their return were tracked by survey IDs. Counts of all incoming forms were updated as they were received. All these documents were optically scanned and edited. Surveys completed in ink were key entered[5]. Scanned survey questions with multiple answers were checked to ensure that the multiple answers were not due to a scanning error (i.e., the scanner erroneously picked up an erased answer as a response). Ten percent of comments were keyed verbatim with expletives deleted. All "other-specify" responses were keyed verbatim.

Throughout the administration of HCSDB, returned surveys were tracked in the mailing system and returns files as surveys were returned, mail was returned PND, and information was received by fax or telephone. A final disposition variable (FLAG_FIN) was developed to classify incoming surveys, and to classify cases where the sample member did not return a survey. The disposition values and outcomes were:

---

[5] All data captured via keying is keyed and verified, yielding an accuracy rate of 99.6%.

- Returned Survey - survey was completed and returned (FLAG_FIN=1);

- Returned Blank (deceased) - survey was returned blank with information accompanying the survey that the sample member was deceased (FLAG_FIN=2);

- Returned Blank (ill, hospitalized, incarcerated) - survey was returned blank with information accompanying the survey that the sample member was ill, hospitalized, or incarcerated (FLAG_FIN=3);

- Returned Blank (all other reasons, e.g., divorced, left service) - survey was returned blank with information that the sample member was divorced, had left the service, or another reason (FLAG_FIN=4);

- Returned Blank (no reason) - survey was returned blank without an explanation (FLAG_FIN=5);

- No Return (no reason) - survey was not returned and no reason was given by sample member (FLAG_FIN=6);

- No Return (deceased) - survey was not returned, sample member deceased (FLAG_FIN=7);

- No Return (ill, hospitalized, incarcerated) - survey was not returned, sample member was ill, hospitalized, or incarcerated (FLAG_FIN=8);

- No Return (active refuser) - survey was not returned, sample member refused to take part in the survey (FLAG_FIN=9);

- No Return (all other reasons) - survey was not returned, another reason provided (e.g. divorce, separation from the military) (FLAG_FIN=10);

- PND (no address remaining) - all addresses were attempted and returned PND (FLAG_FIN=11);

- PND (address remaining at the close of field) - at the close of field the last address used was found invalid, next available was not attempted (FLAG_FIN=12);

- Original Non-Locatable (no address at start of mailing) - substantially incomplete or blank address field prior to the start of the administration of the survey, no mailings attempted (FLAG_FIN=13);

- Return (deceased) - survey was returned completed (at least partially) with information that the sample member was deceased (FLAG_FIN=14);

- Return (incarcerated) - survey was returned completed (at least partially) with information that the sample member was incarcerated (FLAG_FIN=15);

- Return (gender mismatch) - survey was returned completed with information derived from the survey which did not match the intended sample member's gender (FLAG_FIN=20).

Table 4 documents the final disposition data of the survey sample by each beneficiary group as recorded in the system used for mailing. Some sample members did not return a survey and provided a reason why the survey was not returned (i.e., FLAG_FIN values of 9-12). Sample members provided this information through various sources, including collect and 800-number calls, faxes, and letters.

Table 4.
Frequency (n) and percent distribution of final disposition of survey sample by beneficiary group[1]

Final Survey Disposition[2]	Active Duty Personnel	Family Members of Active Duty Personnel	Retirees Under Age 65	Retirees Age 65 or Over	Family Members Age 18-64 of Retirees and Survivors Age 18-64	Family Members Age 65 or Over of Retirees and Survivors Age 65 or Over	Total
Returned Survey[3]	17,154 44.89%	14,096 45.88%	15,096 67.98%	13,243 77.24%	16,689 57.13%	13,423 69.42%	89,701 57.19%
Returned Blank (Deceased)	2 0.01%	6 0.02%	29 0.13%	157 0.92%	29 0.10%	165 0.85%	388 0.25%
Returned Blank (Ill, hospitalized, incarcerated)	0 0.00%	2 0.01%	11 0.05%	42 0.24%	16 0.05%	67 0.35%	138 0.09%
Returned Blank (other reason, e.g. divorced or left service)	50 0.13%	78 0.25%	34 0.15%	66 0.38%	112 0.38%	152 0.79%	492 0.31%
Returned Blank (No reason)	17 0.04%	32 0.10%	17 0.08%	33 0.19%	45 0.15%	64 0.33%	208 0.13%
No Return (No reason)	18,918 49.51%	15,519 50.51%	6,491 29.23%	2,805 16.36%	11,249 38.51%	4,151 21.47%	59,133 37.70%
No Return (Deceased)	18 0.05%	10 0.03%	55 0.25%	326 1.90%	61 0.21%	337 1.74%	807 0.51%
No return (ill, hospitalized, incarcerated)	4 0.01%	2 0.01%	8 0.04%	30 0.17%	16 0.05%	41 0.21%	101 0.06%
No return (Active refuser)	44 0.12%	56 0.18%	16 0.07%	26 0.15%	35 0.12%	65 0.34%	242 0.15%

Table 4. (continued)

Frequency (n) and percent distribution of final disposition of survey sample by beneficiary group

Final Survey Disposition[2]	Active Duty Personnel	Family Members of Active Duty Personnel	Retirees Under Age 65	Retirees Age 65 or Over	Family Members Age 18-64 of Retirees and Survivors Age 18-64	Family Members Age 65 or Over of Retirees and Survivors Age 65 or Over	Total
No Return (all other reasons, e.g., divorced or left service)	78 0.20%	61 0.20%	5 0.02%	9 0.05%	25 0.09%	24 0.12%	202 0.13%
PND (No address remaining)	636 1.66%	459 1.49%	131 0.59%	199 1.16%	508 1.74%	381 1.97%	2,314 1.48%
PND (Address remaining at close of field)	508 1.33%	279 0.91%	86 0.39%	36 0.21%	187 0.64%	44 0.23%	1,140 0.73%
Non-Locatable (No address at start of mailing)	180 0.47%	4 0.01%	26 0.12%	61 0.36%	73 0.25%	307 1.59%	651 0.42%
Returned (Deceased)	0 0.00%	2 0.01%	4 0.02%	13 0.08%	2 0.01%	7 0.04%	28 0.02%
Returned (Incarcerated)	1 0.00%	0 0.00%	1 0.00%	0 0.00%	0 0.00%	0 0.00%	2 0.00%
Returned (Gender mismatch)	604 1.58%	119 0.39%	195 0.88%	99 0.58%	166 0.57%	108 0.56%	1,291 0.82%
Total	38,214 24.37%	30,725 19.59%	22,205 14.16%	17,145 10.93%	29,213 18.63%	19,336 12.33%	156,838 100.00%

[1] Taken from BGCSMPL.

[2] Taken from FLAG_FIN.

[3] Only surveys in this category were included in the 1996 HCSDB data file.

Inclusion Criteria for Surveys in the 1996 HCSDB Data File

Both variables FLAG_FIN and FLAG_DUP were created in advance of the production of the final data file. A survey wave indicator (FLAG_DUP) was developed to identify sample members who returned more than one survey. Each survey was examined to determine whether the survey was from the first wave mailing or the second wave mailing. The data in Table 5 presents the final disposition for all incoming surveys.

Table 5.
Survey wave indicator[1] by final disposition[2]

Final Disposition	Wave 1	Wave 2	Wave 1 (both Wave 1 and Wave 2 were returned)	Wave 2 (both Wave 1 and Wave 2 were returned)	Total
Returned Survey	71,753 79.78%	17,569 19.53%	322 0.36%	295 0.33%	89,939 97.04%
Returned Blank (Deceased)	214 53.37%	162 40.40%	12 2.99%	13 3.24%	401 0.43%
Returned Blank (Ill, hospitalized, etc.)	82 57.75%	54 38.03%	2 1.41%	4 2.82%	142 0.15%
Returned Blank (other reasons)	185 33.39%	288 51.99%	27 4.87%	54 9.75%	554 0.60%
Returned Blank (No reason)	102 35.92%	91 32.04%	43 15.14%	48 16.90%	284 0.31%
Returned (Deceased)	23 74.19%	4 12.90%	1 3.23%	3 9.68%	31 0.03%
Returned (Incarcerated)	2 100.00%	0 0.00%	0 0.00%	0 0.00%	2 0.00%
Returned (Gender mismatch)	981 73.81%	306 23.02%	25 1.88%	17 1.28%	1,329 1.43%
Total	73,342 79.13%	18,474 19.93%	432 0.47%	434 0.47%	92,682 100.00%

Note: This table was generated with data obtained prior to removal of any records from the file.
[1] Taken from FLAG_DUP.
[2] Taken from FLAG_FIN.

        In order to minimize additional processing during the analysis and report generating phase, several types of records were removed from the 1996 HCSDB. Several steps were taken to remove additional surveys from the same sample member, surveys returned from sample members who were identified as deceased, and blank surveys. If a sample member returned more than one survey, then the first wave survey was accepted and placed in the file. If the first wave survey was returned blank, then the second wave survey was selected. During the operations phase, care was taken so that deceased sample members did not receive a survey. If a sample member was very ill or was deceased, someone would call, fax, or mail information to the survey operations center with this information. In some cases, a family member would return a survey and indicate that the person was deceased. In all cases, surveys from deceased sample members were not included in the 1996 HCSDB. In addition to these types of records, blank surveys were also removed. A variable created during the operations phase (MISS_9) that summed all values of -9 (SAS[®]: .) for all original variables was used to identify completely blank surveys. If MISS_9 was equal to 279, then that record was removed from the file.

A final step was taken to ensure that the person completing and returning the survey was the same person the survey was mailed to. A survey record was flagged if the gender of the sample member was different than the self-reported gender. Based on an examination of these cases, it was evident that someone other than the intended sample member completed the survey. In some cases, the self-reported gender in the survey was specified as "female" and the name and gender in the mailing system were "male." In other cases, the self-reported gender in the survey was specified as "male" and the name and gender in the mailing system were "female." Overall, 1,329 records indicating that someone other than the sample member completed the survey were removed from the 1996 HCSDB data file.

Of the 92,682 returned surveys, a total of 2,981 surveys were removed from the 1996 HCSDB data file, yielding 89,701 surveys. These 2,981 surveys included surveys returned blank, surveys from deceased, incarcerated, or ill sample members, multiple surveys from the same respondent, and surveys not completed by the intended recipient. It is possible that a removed survey record may meet one, some, or all of the criteria for exclusion.

Guide to Using the Codebook

The codebook documents all of the variables included in the 1996 HCSDB, as well as describing how variables are coded, and new variables are created. The annotated questionnaire in Appendix A, provides a name for each variable, the values for various response options as they appear on the survey, and coding "notes." Coding notes appear on pages 4-7, 10-11, 15-18, 22, and 23 of the questionnaire. These coding "notes" direct the analyst to refer to various tables in the coding scheme (Appendix B) to determine how selected variables were coded. The annotated questionnaire and coding scheme, when used in conjunction with the data element pages (Appendix E), serve as documentation of data coding and editing of the questionnaire. This guide is an essential tool for the analyst who plans to conduct analyses of the data. For the 1996 survey, coding tables were implemented to provide clear specifications of data coding.

The variables in the data file and the codebook are grouped by logical categories. Pages E-1 through E-312 (in Appendix E) include variables derived from the sample which are used for stratification and classification, variables derived from the DEERS files, survey item variables, and variables used in the coding tables. The survey items begin on page E-28 and continue through E-298. Constructed variables which will be utilized in various reports begin on page E-313.

Variable Naming Conventions

The naming conventions of the variables original variables used for items in this survey capture the type of survey, the year of the survey and the questionnaire item. For example, question 1 is equivalent to the variable H9601. The first letter "H," indicates the type of survey (i.e., health care for beneficiaries 18 years old or older), and "96" indicates the year of the survey. The information following the year, "01" indicates the survey question. For mark all questions, each response is treated as a separate variable in the coding scheme.

Other naming conventions begin with "SR" (mnemonic for self-report), and are typically used for demographic questions. This includes variables such as gender (SRSEX; question 21), age (SRAGE; question 78), marital status (SRMARST; question 79), highest level of education (SRED; question 80), ethnic group (SRHISPA; question 81), race (SRRACE; question 82), and military facility (SRMMTF; question 83). The variables that appear after page E-312 have naming conventions to set them apart from the original survey variables. These constructed variables include case weights, and over 100 additional variables required by the Office of the Assistant Secretary of Defense (Health Affairs).

Variable Coding Guidelines

The guiding premise of coding the HCSDB is to clearly communicate all aspects of coding of the questionnaire items so that a wide range of analysts can access and analyze the data. Datasets such as this one are analyzed repeatedly over time by different people within the government and outside the government. The coding scheme and the annotated questionnaire that accompanies the coding scheme were based on this guiding premise. Based on experiences with the 1994-95 HCSDB, several steps were taken to develop a coding scheme that preserves original data and creates new variables based on specific response patterns within the survey.

Coding of response options on the annotated questionnaire. As each survey is read by a scanner, the scanner produces a file that is converted to a code for each response option (bubble) on the questionnaire. The annotated questionnaire (Appendix A), presents codes for each response option. Most of the response options are assigned a value of 1, 2, 3, 4, 5, etc. An exception to this rule are cases where "0" is a true value (e.g., questions 19, 41, and 54). Missing data for survey questions are coded -9 (SAS®: .), and multiple response errors (questions where a respondent marked more than one response to a question) are coded -8 (SAS®: .A). The analyst will also find values of -6 (SAS®: .N) for questionnaire items where "not applicable" is a valid response option. These conditions are described in Table B-1 in Appendix B.

Coding of special types of response options. In addition to the values presented in the annotated questionnaire, there are other special codes to consider (see Table B-1 in Appendix B). The use of "-6 (SAS®: .N) is not only used to indicate a "not applicable" response, but is also used to indicate a "valid skip." These are conditions where the respondent answered a question and followed instructions to skip to another question in the survey. The coding of items skipped are typically assigned values of -6 (SAS®: .N). In many cases, additional "valid skip" codes are used to capture skipping out across several items in the survey. Table B-2 presents nine additional types of special codes. An example of how this "skipping out" across items is coded, can be described using question 38 as an example. If a respondent answers "no" to question 38 (indicating no military health care), and leaves questions 39 through questions 47 all blank, then questions 39 through 47 are coded as -64 (SAS®: .T), "not applicable, no military health care."

The analyst will find that these types of special codes are treated as missing in a standard SAS® frequency distribution, unless the "missing" option is added to the SAS® frequency procedure. In general, an analyst would want to use only those cases of interest, so the treatment

of the special types of not applicable response options as types of "missing" are generally preferred. Special value codes will often be recoded or reformatted by the data analyst to a one value so that tabulations/ percentages would represent only those to whom the question applied.

   Coding of skip (screening) questions. In general, most people who respond to a mail survey follow instructions for skipping out of questions, and do so appropriately. However, there are cases where a respondent fails to follow the instructions to skip out of a question. Even though the frequency of these cases is small, it is essential that all patterns be accounted for in the coding scheme. For the 1996 HCSDB, care was taken to create a recoded version of the original skip question to account for cases where a respondent skipped out appropriately or not. For each question where the respondent is instructed to skip one or more questions, a recoded version of the skip question appears on the annotated questionnaire (suffixed in "R") in addition to the original skip question variable. Examples of this include the screening question on smoking (H9616, H9616R), the screening question on a particular place for health care (H9627, H9627R), and the screening question on private medical insurance (H9635, H9635R). When examining items presented in the annotated questionnaire, it is recommended that the analyst use recoded versions of the original variables (suffixed in "R"). However, there are conditions when the use of the recoded versions (suffixed in "R") may yield an increase in the percent of "yes" responses. The analyst should carefully examine the frequencies of H9635R with H9635, H9638R with H9638, and H9651 with H9651R when deciding whether to use the original or the recoded version.

   Backward and forward coding. The coding scheme and coding tables present all possible combinations of responses to a skip (screening) question and data coding. The variables suffixed in "R" have been created using "backward" coding and/or "forward" coding. With backward coding, if a person marks an answer on the starting question of a skip pattern that indicates that they should skip out of items but they go ahead and mark those items, a recoded version of that skip question would be created. This is in contrast to forward coding where the data on the starting question is accepted as marked and all data for the remaining items is ignored. In the case of forward coding, a recoded version of the skip question and a recoded version of each subsequent question in the series is created. Both the original version of the variable and the recoded version of each variable are retained on the file. Forward coding is typically used in cases where the true gender of the person indicates that they should have skipped out of the series of the items. For those cases, the items in the series are coded as not applicable, "-6 (SAS®: .N)." Forward coding is best utilized when there are three response options for the screening question (e.g., current smoker, ex-smoker, or never smoked). When three response options exist, it is sometimes difficult to reconcile what the respondent intended. When two response options exist on a skip question (e.g., yes, no), backward coding is typically used.

   Coding of "mark-all-that-apply" questions. DMDC treats mark-all-that-apply items as if they were a series of yes/no items and codes them as individual variables with codes of "1, marked" and "2, not marked" similar to codes of "1, yes" and "2, no." For cases where all of the mark-all-that-apply items are left blank, all of the response options for the question are coded as a -9 (SAS®: .). There are cases where a "not applicable" response option appears as one of several response options for a mark-all-that-apply question. The responses to the question are

treated as if they contained a skip pattern. That is, if the "not applicable" response is marked and all other responses are blank, then the "not applicable" response is coded as a "1, marked", and all other responses are coded as "not applicable" -6 (SAS®: .N). On the other hand, if the "not applicable" response is marked and any other response is marked, then the "none of the above" response option is treated as a multiple response error -8 (SAS®: .A), and the other response options are coded as "1, marked" or "2, not marked." There are two mark-all-that-apply questions in the 1996 HCSDB, question 49 and question 68.

Example of a Coding Table

Figure 1 provides an example of a coding table for questions 35 and 36. This example has been provided so that an analyst will have an understanding of how some sections of the questionnaire have been coded. Each row of the table provides specifications on all possible coding combinations represented in various columns. The variable identified in the first column appears on the data set with values indicating all possible combinations of coding. It was retained on the data file so that the analyst would have the option of examining combinations. The second column presents all values of the skip (screening) question. The third column presents all possible values of the question following the skip question -- either blank or marked. When several questions follow a skip question, the values are either "all blank" or "at least one marked" are indicated. A mark can be a valid response from the annotated questionnaire or a multiple response condition -8 (SAS®: .A) where more than one response was marked on the questionnaire. The fourth column provides the values of the recoded version of the original skip question. The fifth column shows any coding on the question (or questions) following the skip question. The last column is a note indicating if a backward coding or forward coding operation was implemented, with "B" indicating backward coding and "F" indicating forward coding.

Specifically, Figure 1 walks the analyst through selected rows and columns of a coding table. For the first row (where N5 is "1"), if H9635 is "yes" and H9636 is "marked" then H9635R is coded as "yes" and the value of H9636 stands as it appears on the annotated questionnaire. To understand one type of backward coding, the analyst should examine the third row (where N5 is equal to "3"). Under this coding condition, if H9635 is "no" and H9636 is "marked" then H9635R is coded as "yes" and the value of H9636 stands as it appears on the annotated questionnaire. [For H9635R, the percent of "yes" responses are higher than the original version due to this specific backward cleaning step]. Remember that backward cleaning considers the responses following the skip (screening) question as valid responses and adjusts the skip question accordingly (in this case provides a different value for the recoded screening question). Another type of backward coding occurs when a respondent leaves the skip (screening) question blank. This scenario is demonstrated on the fifth row of the table (where N5 is equal to "5"). Under this condition, if H9635 is blank and H9636 is "marked" then H9635R is coded as "yes" and the value of H9636 stands as it appears on the annotated questionnaire. The analyst should carefully examine the frequencies of H9635R with H9635 when deciding whether to use the recoded or the original version.

Figure 1. Example of A Coding Table

Coding Table for Note 5:
H9635, H9636, H9635R

N5	H9635 is:	H9636 is:	H9635R is coded as:	H9636:	
1	1 yes	mrkd	1 yes	stands	
2	1 yes	blank	1 yes	is coded as blank	
3	2 no	mrkd	1 yes	stands	B
4	2 no	blank	2 no	is coded as .N, NA	
5	blank	mrkd	1 yes	stands	B
6	blank	blank	blank	is coded as blank	
7	mre	mrkd	1 yes	stands	B
8	mre	blank	mre	is coded as blank	

Example of a Codebook Page

Figure 2 provides an example of the presentation of a survey item in the codebook. Data element pages in Appendix E provide the question as it appears on the questionnaire, the variable name, the frequency distribution of response options as they are written on the survey, missing values, and special types of "not applicable" values. Where appropriate, additional text regarding the question or the coding of the question is presented at the bottom of each page. Annotations include an indication of those response distributions too lengthy to be presented in the codebook tables, warnings when percentile totals do not sum to 100% due to rounding within individual cells, and descriptive text provided from the coding notes. The illustration is explained and references to page elements are provided in the text that follows.

Figure 2. Example of A Codebook Page

[1]The 1996 Health Care Survey of DoD Beneficiaries

[2]Which of the following explains why you did NOT get most of your medical care at a military facility in the past 12 months? MARK ALL THAT APPLY.

[3]H9649A- [4]The military facility is too far away

[5]FREQ	[6]PERCENT	[7]OS VALUE	[8]SAS VALUE	[9]MEANING
3898	4.3	-9	.	All blank
42906	47.8	-6	.N	Not applicable or valid skip
11607	12.9	1	1	Marked
31290	34.9	2	2	Not marked
[10]89701	[11]99.9	TOTALS		

[12]IF NONE OF THE RESPONSE BUBBLES FOR THE MARK ALL QUESTION WERE MARKED (EXCLUDING THOSE WHO SKIPPED OUT OF H9648) THEN ALL ITEMS IN THE SERIES WERE TREATED AS MISSING.

Explanation of Figure 1:
Contents Of A Codebook Page For A Survey Item

1. Codebook title.

2. Survey question or text identifying the variable.

3. Variable name.

4. Response option (generally applicable to mark-all-that-apply questions).

5. Frequency (count) of responses, including missing values and special codes.

6. Percentage of total responses represented by each value, including missing values and special codes.

7. Response values stored in flat file (OS file).

8. Response values stored in SAS® format.

9. Explanation of the response value codes. Special codes were used to denote the various types of missing values, as well as special skip pattern values.

10. Total frequency (count).

11. Total percent.

12. Text helpful to analyst and/or descriptive text such as "Too Numerous to tabulate."

REFERENCES

Chu, Flores-Cervantes, and Latta (1996). *Sample Selection Procedures for the 1996 Health Care Survey of DoD Beneficiaries (DMDC Study Report 95-012)*. Arlington, VA: Defense Manpower Data Center.